

Toetsing in context: een onderzoek naar de effecten van klinische context in een anatomietoets

R. Dijcks, G.M. Verwijnen, L.W.T. Schuwirth, H. van Mameren, C.J.A.H. Prince, A.J.J.A. Scherpbier, C.P.M. van der Vleuten

Samenvatting

Inleiding: Volgens resultaten van cognitief psychologisch onderzoek moeten onderwijs en toetsing zoveel mogelijk in de context van de echte praktijk worden aangeboden. Het is de vraag of het gebruik van partiële context bij kennistoetsen zinvol is. Om dit laatste te onderzoeken is een antwoord gezocht op de vraag: welk effect heeft partiële context bij een anatomietoets op toetsresultaten en psychometrische variabelen?

Methode: Twee inhoudelijk identieke versies van een anatomietoets (A-toets), één met en één zonder context, zijn voorgelegd aan twee at random geselecteerde groepen van in totaal 58 vierdejaars geneeskundestudenten. Alle studenten maakten ook dezelfde contextloze B-toets (48 anatomievragen) en vulden een vragenlijst in over de A-toets.

Resultaten: De onderzoeksgroepen verschillen niet significant in de scores op de B-toets ($t=-1.1167$, $df=56$, $p=.248$) en de A-toets ($t=-.537$, $df=56$, $p=.593$). De A-toetsen verschillen niet in betrouwbaarheid over items en over casus. De B- en de A-toets correleren sterk ($R=1$ (A zonder context) en $R=.96$ (A met context)). De Pearsoncorrelatie tussen de p-waarden van de A-toetsversies bedraagt .939. De meningen van de studenten over de verschillende toetsversies verschillen niet significant.

Beschouwing: De onderzoeksgroepen zijn vergelijkbaar. Er is geen verschil gevonden tussen de effecten van de A-toets met en zonder context. Kennelijk heeft partiële context bij deze kennisvragen geen toegevoegde waarde. Verder onderzoek is nodig om na te gaan of dit toegeschreven kan worden aan de aard van de context of aan het onderwijs waardoor de context onvoldoende betekenis heeft voor studenten. Beide conclusies hebben belangrijke implicaties voor het onderwijs. (Dijcks R, Verwijnen GM, Schuwirth LWT, Mameren H van, Prince CJA, Scherpbier AJJA, Vleuten CPM van der. Toetsing in context: een onderzoek naar de effecten van klinische context in een anatomietoets. Tijdschrift voor Medisch Onderwijs 2002;21(6): 259-266.)

Inleiding

Uit cognitief-psychologisch onderzoek is gebleken dat het verwerven van expertise verloopt via de ontwikkeling en uitbreiding van een kennisnetwerk.^{1 2} Dit netwerk bestaat uit begrippen en voorbeelden die met elkaar verbonden zijn. Nieuwe informatie activeert reeds aanwezige begrippen en voorbeelden, waardoor vervolgens weer andere begrippen geactiveerd worden. Door informatie goed op te slaan in het geheugen wordt een goed functionerend netwerk opgebouwd, dat

ertoe leidt dat informatie sneller en beter beschikbaar is. Voor een juiste opslag in het geheugen is het van belang om betekenis te geven aan de leerstof. Daarom verdient het aanbeveling de leersituatie zoveel mogelijk in overeenstemming te brengen met de praktijksituatie waarin het geleerde toegepast moet worden.¹ Leren dient dus bij voorkeur plaats te vinden in een relevante context.

Wanneer studenten leren in context, is het logisch om ook in context te toetsen. De klinische details en de contextuele

informatie over de patiënt vormen belangrijke triggers die de in het kennisnetwerk aanwezige relevante voorkennis kunnen activeren. Studenten zullen zo beter en sneller de relevante kennis kunnen (re)produceren.

De mate van authenticiteit van de situatie waarin geleerd en getoetst wordt, kan variëren. Zo kan er onderscheid gemaakt worden tussen partiële context en rijke context. Recent onderzoek heeft effecten aangetoond van toetsing in rijke context.³ Uit onderzoek van Schuwirth et al. bleek dat vragen in een korte casus andere denkprocessen oproepen dan feitenkennisvragen en dat deze denkprocessen meer gericht zijn op probleemoplossen. In het onderzoek van Schuwirth was echter sprake van een eenheid van casus en vraag: de vraag was niet zonder de casuïstiek te beantwoorden.^{3 4}

Bij partiële context bij kennistoetsen wordt niet voldaan aan deze voorwaarde. De informatie in de casus biedt een context die meer illustratief is dan noodzakelijk om de vraag te kunnen beantwoorden. Het is de vraag of partiële context zinvol is. Het in dit artikel gepresenteerde onderzoek is opgezet om de effecten op toetsresultaten te meten van partiële context bij een kennistoets. Om deze effecten te meten zijn vergelijkingen gemaakt tussen de resultaten op anatomiekennistoetsen met en zonder context. De vraagstelling van het onderzoek luidt: welk effect heeft partiële context bij een anatomietoets op toetsresultaten en psychometrische variabelen?

Methodie

Om de effecten van partiële context bij toetsvragen te meten zijn de resultaten van inhoudelijk identieke anatomietoetsen, één zonder casus en één met casus, geanalyseerd en vergeleken. Hierbij is gekeken

naar de toetsscore, de betrouwbaarheid, de correlatie tussen de p-waarden en de mening van studenten over de toets. De uitgangspunten hierbij waren de volgende: als de toegevoegde casuïstiek een verrijking van de stimulus vormt waardoor het kennisnetwerk beter geactiveerd wordt, zal de score op de vragen met context hoger zijn. Als context niet van belang is voor de beantwoording van de vragen, zou context een vorm van ruis kunnen introduceren, die de betrouwbaarheid van de vragen vermindert. Omdat de casus de relevantie van de bijbehorende items illustreert, is het aannemelijk dat de correlaties tussen de items van dezelfde casus hoger zijn dan de correlaties tussen items van verschillende casus. Dit betekent dat de berekening van de betrouwbaarheid over de items zou leiden tot een overschatting van de betrouwbaarheid. Met andere woorden: als er een effect is van de casus, zal de betrouwbaarheid gemeten over de items hoger zijn dan de betrouwbaarheid gemeten over de casus. Als casuïstiek leidt tot andere denkprocessen dan contextloze vragen, zou het effect hiervan blijken uit een lage (ware) correlatie tussen de p-waarden van de toetsitems met en zonder casus. Als de casus niet leidt tot andere denkprocessen, kan het nog steeds zo zijn dat items met context gemiddeld makkelijker zijn, hetgeen zou kunnen blijken uit verschillen in p-waarden per item. Ten slotte is de mening van de studenten over de toets onderzocht, omdat verwacht werd dat studenten de versie met context relevanter, leerzamer en leuker zouden vinden; het is immers een betere weergave van de praktijk.

Proefpersonen

Alle vierdejaars medische studenten van de Universiteit Maastricht werden benaderd. De toets werd afgenomen in het

laatste onderwijsblok van het vierde jaar. De studenten hadden dus nog geen ervaring met co-assistentenschappen. Deelname aan de toets leverde vijftig gulden op. De deelnemers werden per geslacht at random toebedeeld aan twee groepen. Hoewel de groepen gerandomiseerd waren, is ook nog het kennisniveau van beide groepen geanalyseerd door de resultaten van de deelnemende studenten op drie reguliere voortgangstoetsen te vergelijken. Omdat bekend is dat vrijwilligers vaak betere studenten zijn, werden aan de hand van de reguliere toetsuitslagen ook de resultaten van de deelnemende studenten vergeleken met die van de rest van de jaargroep. Daarnaast werden de toetsscores van de 48 losse anatomievragen (B-subtoets), die voor beide groepen gelijk waren, vergeleken.

Instrument

Voor dit onderzoek is een bestaande anatomietoets aangepast.⁵ Deze toets bestond oorspronkelijk uit een subtoets A met partiële context in de vorm van korte casus en een contextloze subtoets B. De inhoud van beide toetsen was verschillend. Voor dit onderzoek is subtoets A aangepast en subtoets B ongewijzigd overgenomen.

Subtoets A

Van subtoets A werden twee versies geconstrueerd: een contextgebonden versie (vragen gekoppeld aan casus) en een contextloze versie (dezelfde vragen zonder casus).

De contextgebonden versie bestond uit vijftien casus met 129 open, meerkeuze- of juist/onjuistvragen. Hiervan is een contextloze versie gemaakt bestaand uit 129 losse vragen. In de contextloze versie is alle patiënteninformatie weggelaten. Alleen de informatie die noodzakelijk is voor beantwoording van de vragen is ge-

handhaafd. In beide versies was de vraagvolgorde gelijk. Er was geen 'weet-niet-optie'. Voorbeelden van beide versies worden gegeven in figuur 1.

Subtoets B

Deze subtoets bestond uit 48 juist/onjuistvragen met betrekking tot anatomiekennis afkomstig uit bestaande Maastrichtse voortgangstoetsen.⁵

Vragenlijst

Om de waardering van studenten te meten werd een vragenlijst met beweringen over onder andere moeilijkheidsgraad, gokpercentage en relevantie van toets A afgenomen. Het oordeel over deze beweringen moest ingevuld worden op een vijfpuntsschaal (1=helemaal oneens; 5=helemaal eens).

Afname en verwerking

De studenten uit groep I maakten de contextloze versie van subtoets A en de studenten uit groep II maakten de contextgebonden versie van subtoets A. De studenten uit beide groepen maakten daarnaast subtoets B.

De gehele toets werd per computer afgenomen en nam twee uur in beslag. De toets begon voor beide groepen met de 48 vragen van subtoets B, die in dezelfde volgorde werden gesteld. De vraagvolgorde van beide subtoetsen A was ook hetzelfde voor beide groepen. Het was niet mogelijk om terug te bladeren in de toets en een gegeven antwoord later te corrigeren. Afname vond plaats onder standaard-examencondities. De vragenlijst werd door alle studenten na de toetsafname ingevuld.

Na afname werden de open vragen handmatig gescoord volgens een van tevoren ontwikkelde antwoordsleutel. Deze was opgesteld door twee anatomen. Bij twijfel werden alternatieve benamingen

Figuur 1. Voorbeelden uit de contextgebonden en de contextloze versie van subtoets A.**Contextgebonden versie**

Mevrouw Visser, een 49-jarige huisvrouw, komt bij de huisarts. 's Morgens, onder de douche, voelde ze een harde, pijnloze zwelling in haar rechterborst. Een tante is overleden aan borstkanker, ze is bang dat haar nu hetzelfde zal overkomen. Er is geen tepeluitvloed. Ze heeft geen andere klachten en is niet afgevallen. Ze heeft 3 kinderen, die allen borstvoeding hebben gehad. Ze heeft nooit ontstekingen van haar borsten gehad. Bij lichamelijk onderzoek ziet de huisarts dat de rechter tepel iets hoger staat dan de linker.

In het buitenste bovenste kwadrant van de rechter borst is een lichte intrekking zichtbaar.

De huisarts voelt daar een ca. 2 cm grote zwelling, hard aanvoelend, mogelijk vastzittend aan de onderlaag.

Vraag 1.

Een tumor in een borst kan doorgroeien in de spierlaag tussen borstweefsel en thoraxwand.

- Welke spier ligt het meest oppervlakkig in deze spierlaag?

Vraag 2.

De huisarts wil onderzoeken of de tumor vastzit aan de m. pectoralis major. Bij dit deel van het onderzoek moet deze spier aangespannen zijn. Hij laat mevrouw Visser haar rechter schouder optrekken (elevatie).

Tijdens deze beweging contraheert de m. pectoralis major.

- Deze stelling is juist/onjuist?

Contextloze versie**Vraag 1.**

Een spier ligt het meest oppervlakkig in de spierlaag tussen mammaweefsel en thoraxwand.

- Welke spier is dit?

Vraag 2.

Tijdens het optrekken van de schouder (elevatie) contraheert de m. pectoralis major.

- Deze stelling is juist/onjuist?

aan hen voorgelegd. Na analyse ontvingen de studenten een schriftelijke evaluatie met hun eigen score en de gemiddelde score van de jaargroep.

Data-analyse

Vooraf werd een poweranalyse verricht. Er bleken twintig personen per groep nodig te zijn om een scoreverschil ter grootte van 5% te kunnen detecteren ($\alpha = 0.05$, $\beta = 0.10$).⁶ De gemiddelde procentuele scores en de standaarddeviaties (SD) werden berekend. De resultaten van beide groepen werden vergeleken met behulp

van de onafhankelijke t-test. De betrouwbaarheid (Cronbachs alfa voor interne consistentie) van beide toetsen en toetsversies werd berekend. De betrouwbaarheid is zowel met items als met casus als toetseenheid berekend. De Pearsoncorrelatie tussen de scores van de A-toets en de B-toets werd berekend voor beide groepen deelnemers. Er vond een correctie voor attenuatie plaats. In een itemanalyse werden de p-waarden bepaald. De correlatie tussen de p-waarden van de contextloze versie en de contextgebonden versie is berekend. De verschillen in waardering

van de toets tussen beide groepen studenten werden op significantie getoetst met een Mann-Whitneytoets.

Resultaten

Proefpersonen

Het aantal deelnemers bedroeg 58 (28% van de jaargroep) waarvan 30 studenten in groep I en 28 in groep II. De resultaten van drie reguliere voortgangstoetsen vertonen geen significante verschillen tussen beide groepen of tussen de deelnemers aan de toets en hun niet-deelnemende jaargenoten. De deelnemers vormen dus een representatieve afspiegeling van hun jaargroep. De verschillen tussen de gemiddelde procentuele scores van de B-subtoets van beide groepen (tabel 1) zijn statistisch niet significant (onafhankelijke t-test: $t = -1.167$, $df = 56$, $p = .248$).

Toetsen

De toetsparameters zijn weergegeven in tabel 1. De gemiddelde procentuele score, SD en Cronbachs alfa van de A-subtoets zijn per versie berekend. Het verschil tussen de scores op de verschillende versies is statistisch niet significant (onafhankelijke t-test: $t = -.537$, $df = 56$, $p = .593$). Voor subtoets A is de betrouwbaarheid per versie berekend zowel met items als toets-eenheid ($\alpha = .89$ voor groep I en $.88$ voor

groep II) als met casus als toets-eenheid ($\alpha = .87$ voor groep I en $.88$ voor groep II). De betrouwbaarheid (α) van de B-subtoets is $.55$ voor groep I en $.47$ voor groep II. De Pearsoncorrelatie tussen de B-subtoets en de contextloze en de contextgebonden versie van subtoets A bedraagt respectievelijk $.745$ en $.650$. De waarden voor de ware correlatie (gecorrigeerd voor attenuatie) zijn respectievelijk $r = 1$ en $r = .96$. De Pearsoncorrelatie tussen de p-waarden van de versies van subtoets A is $.939$. Bij bestudering van de p-waarden blijkt er slechts bij 26 van de 129 vragen een behoorlijk verschil te bestaan tussen de contextloze en contextgebonden versie (p -waarde $> .15$): bij 14 vragen scoort de contextloze versie beter en bij 12 vragen de contextgebonden versie. Deze verschillen bleken bij inhoudelijke analyse niet terug te voeren op specifieke vraagkenmerken. Student-assistenten van de vakgroep anatomie/embryologie scoorden in een ander onderzoek hoger dan de overige studenten.⁵ Indien zij worden verwijderd uit het bestand, vallen drie studenten af. Dit heeft geen invloed op de hierboven gerapporteerde bevindingen.

Vragenlijst

De scores op de vragenlijst betreffende de waardering van studenten voor de verschillende versies van subtoets A zijn sa-

Tabel 1. Toetsresultaten.

	Toets A contextloos			Toets A contextgebonden			Toets B		
	Score (SD)	Alfa (item)	Alfa (casus)	Score (SD)	Alfa (item)	Alfa (casus)	Score (SD)	Alfa	Correlatie A:B
Groep I (N=28)	60.3 (11.1)	.89	.87				62.4 (9.8)	.55	.745 (R=1)
Groep II (N=30)				61.8 (9.8)	.88	.88	65.3 (8.8)	.47	.650 (R=.96)

mengevat in tabel 2. Er zijn geen significante verschillen gevonden. Wel is er een lichte tendens om de contextgebonden versie hoger te waarderen. De studenten geven aan de toets relevant, leuk en leerzaam te vinden. Ook geven veel studenten aan gekokt te hebben.

Beschouwing

In dit onderzoek zijn geen verschillen gevonden tussen de resultaten van twee versies van een anatomietoets, die alleen verschillen in de aan- of afwezigheid van context in de vorm van een korte casus. De aanwezigheid van partiële context heeft in dit onderzoek geen effect op de toetsprestatie. Ook de betrouwbaarheid van beide versies van de A-subtoets is nagenoeg gelijk. Bestudering van de p-waarden per vraag geeft hetzelfde beeld. Slechts bij enkele vragen is er sprake van een behoorlijk verschil, maar dit is niet systematisch in een bepaalde richting en de verschillen kunnen niet teruggevoerd worden op specifieke vraagkenmerken. De Pearsoncorrelatie voor de p-waarden van de verschillen tussen de vragen van de versie met en zonder context is hoog. De context verandert dus ook niets aan de relatieve moeilijkheid van de items. Er is

ook geen significant verschil in de waardering van beide toetsen door studenten. Men vindt beide versies even leerzaam, relevant en leuk. De verschillen op de score per vraag zijn minimaal.

Een mogelijke verklaring voor het ontbreken van verschillen is dat de contextloze versie toch een vorm van context bezit, doordat de vraagvolgorde gelijk is aan die van de versie met context. De versie met casuïstiek moet een bepaalde 'logische' volgorde hebben die past in het ziekteverloop van de patiënt. Als dit zo is, zou verwacht mogen worden dat er een verschil is tussen de alfa's over de itemscores en de casusscores. Bij de versie met partiële context zal de betrouwbaarheid over de casusscores lager zijn dan de betrouwbaarheid over de itemscores. Dat dit niet het geval is, geeft aan dat zelfs de casus geen bindende factor is die de betreffende vragen meer inhoudelijk samenhangend maakt. Desalniettemin zou een vervolgonderzoek met een derde toetsversie waarin de volgorde van de vragen gewijzigd is, zinvol zijn.

Een eventueel contexteffect zou ook overschaduwd kunnen zijn door gokgedrag. Dit kan te maken hebben met het ontbreken van een weet-niet-optie bij de

Tabel 2. De waardering van studenten voor de versies van toets A.

	Waardering toets A*	
	Contextloos Mean (SD)	Contextgebonden Mean (SD)
Ik vond de toets relevant.	3.93 (1.15)	4.00 (1.11)
Ik denk dat ik de toets goed gemaakt heb.	2.36 (1.02)	2.70 (0.79)
Ik heb vaak moeten gokken.	3.46 (1.14)	3.13 (1.04)
Ik vond de toets leuk.	3.75 (1.17)	3.97 (0.93)
Ik vond de toets leerzaam.	3.71 (1.08)	3.97 (1.03)

* op vijfpuntsschaal: 1=helemaal oneens en 5=helemaal eens.

vraagbeantwoording. Uit de respons op de vragenlijst blijkt immers dat veel studenten gegokt hebben. Puur gokgedrag levert een random invulpatroon op dat de gevoeligheid om verschillen tussen toetsen aan te tonen, vermindert. Aan de andere kant blijkt echter uit onderzoek dat het invoeren van een weet-niet-optie ook vaak een negatieve invloed heeft op de betrouwbaarheid. Ook blijkt dat bij 'gokken' vaak partiële kennis gebruikt wordt, en dat maar zelden echt puur gegokt wordt.^{7 8} De gevolgen van een weet-niet-optie zouden in vervolgonderzoek bestudeerd kunnen worden.

Een andere verklaring voor de gevonden resultaten is dat het kennisnetwerk van de student onvoldoende geactiveerd wordt door de context en de student dus niets met de context doet. Hiervoor zijn twee verklaringen mogelijk. De partiële context in de vorm van een casus is niet relevant. Het heeft dan geen zin om casuïstiek in een toets te verwerken, omdat er geen effect waarneembaar zal zijn. Indien men wel context wil toevoegen, moet men zich houden aan duidelijke richtlijnen, met name aan korte en relevante context.^{3 4 9} Het is ook mogelijk dat de studenten nog niet competent genoeg zijn om de toegevoegde waarde van de partiële context te herkennen. Doordat het kennisnetwerk nog niet goed ontwikkeld is, herkent de student de betekenis niet van de mogelijkheden die de context geeft. De kennis is wel in het geheugen aanwezig, maar nog niet toegankelijk voor reproductie. Een mogelijke oorzaak hiervoor is dat het onderwijs onvoldoende in context is aangeboden en dat studenten daarom nooit voldoende betekenis aan de leerstof hebben toegekend. Van groot belang is ook dat studenten in het vierde jaar nog weinig klinische ervaring hebben. Het zou daarom interessant zijn deze studie te

herhalen bij studenten na hun co-assistentenschappen. Wellicht kunnen vierdejaars studenten de toegevoegde waarde van praktijkgegevens nog niet waarderen. Het is bekend dat onder invloed van de praktijk contextuele factoren een steeds belangrijker rol gaan spelen. Experts maken veel meer gebruik van contextuele informatie en genereren daardoor betere diagnostische hypothesen dan basisartsen.¹¹ Een eerder contact met de praktijk gedurende de opleiding kan hier verandering in brengen.

Samenvattend kan het volgende gesteld worden. Er zijn geen significante verschillen waarneembaar tussen een toets met partiële context in de vorm van korte casus en een toets met contextloze vragen voor wat betreft toetsscore, psychometrische eigenschappen en waardering door studenten. Het ligt voor de hand dat studenten oordelen dat de context niet ter zake doet om een antwoord op de vraag te geven en dat zij de context wel lezen, maar niet gebruiken bij het beantwoorden van de vraag. Of dit aan de casus ligt, doordat deze wellicht irrelevant is, of aan de student doordat deze (nog) niet heeft geleerd hoe met contextuele factoren om te gaan, is nog niet helder. Verder onderzoek naar context in toetsing – maar zeker ook naar context in onderwijs – is daarom wenselijk.

Literatuur

1. Regehr G, Norman GR. Issues in cognitive psychology: implications for professional education. *Acad Med* 1996;71(9):988-1001.
2. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise; theory and implications. *Acad Med* 1990;65(10):611-22.
3. Schuwirth LWT, Verheggen MM, Vleuten CPM van der, Boshuizen HPA, Dinant GJ. Do short cases elicit different thinking processes than factual knowledge questions do? *Med Educ* 2001;35:348-56.
4. Schuwirth LWT, Blackmore DE, Mom E, Wildenberg F van den, Stoffers HEJH, Vleuten

- CPM van der. How to write short cases for assessing problem solving skills. *Med Teach* 1999; 21(2):144-50.
5. Prince CJA, Mameren H van, Hylkema N, Drukker J, Scherpbier AJA, Vleuten CPM van der. Does problem based learning lead to deficiencies in basic science knowledge: an empirical case on anatomy. *Med Educ*, accepted.
 6. Schuwirth LWT. De bepaling van de steekproefgrootte: power analyse. *Bulletin Medisch Onderwijs* 1994;13:167-72.
 7. Muijtens AMM, Mameren H van, Hoogenboom RJI, Evers JLH, Vleuten CPM van der. The effect of a 'don't know' option on test scores: number-right and formula scoring compared. *Med Educ* 1999;33:267-75.
 8. Diamond J, Evans W. The correction for guessing. *Reviews in Educational Research* 1973;43:181-91.
 9. Schuwirth LWT, Vleuten CPM van der, Kock CA de, Peperkamp AGW, Donkers HJLM. Computerized case based testing: a modern method to assess clinical decision making. *Med Teach* 1996;18(4):295-300.
 10. Dolmans DHJM, Snellen-Balendong H, Wolfhagen HAP, Vleuten CPM van der. Seven principles of effective case design for a problem based curriculum. *Med Teach* 1997;19(3):185-9.
 11. Hobus PPM, Hofstra ML, Boshuizen HPA, Schmidt HG. De context van de klacht als diagnosticum. *Huisarts Wet* 1988;31:261-3.

De auteurs:

Drs. R. Dijcks, zesdejaars student geneeskunde, student-assistent Skillslab.

G.M. Verwijnen, arts/hoofd Skillslab, Skillslab.

Dr. L.W.T. Schuwirth, arts/medisch onderwijskundige, Capaciteitsgroep Onderwijsontwikkeling & Onderwijsresearch.

Dr. H van Mameren, arts-anatoom, Capaciteitsgroep Anatomie/embryologie.

Drs. C.J.A.H. Prince, huisarts/onderzoeker, Skillslab.

Prof. dr. A.J.J.A. Scherpbier, arts, wetenschappelijk directeur Onderwijsinstituut.

Prof dr. C.P.M. van der Vleuten, voorzitter Capaciteitsgroep Onderwijsontwikkeling & Onderwijsresearch.

Allen zijn verbonden aan de Faculteit der Geneeskunde van de Universiteit Maastricht.

Correspondentieadres:

R. Dijcks, Skillslab, Postbus 616, 6200 MD Maastricht. tel.: 043-3881794, r.dijcks@sk.unimaas.nl.

Summary

Introduction: Research in cognitive psychology has demonstrated the desirability of teaching and testing within a practice-related context. Recent research has suggested, however, that partial context is not effective in knowledge tests. In order to investigate this, we addressed the question: What are the effects of partial context in an anatomy test on test results and psychometric variables.

Method: Two anatomy tests, identical in content but one with and one without context, were each administered to one of two groups of 48 randomly selected fourth year medical students. All students were also given another anatomy test without context (test B) and answered questions about test A.

Results: Test scores yielded no significant differences between the student groups (B test scores: $t=-1.1167$, $df=56$, $p=.248$; A test scores: $t=-.537$, $df=56$, $p=.593$). The A tests showed no differences in reliability across items and across cases. The correlations between the B test and the A test versions are strong ($R=1$ for A without context; $R=.96$ for A with context). The Pearson correlation for the p-values of the A tests is .939. Student opinions did not differ significantly between test versions.

Discussion: The student groups are comparable. No differences were demonstrated between the test versions with and without context. Apparently, partial context has no added value for this type of knowledge test. Further studies will have to determine whether these findings are attributable to context characteristics or to failure of the educational programme to enable students to put knowledge into context. Either conclusion has important educational implications. (Dijcks R, Verwijnen GM, Schuwirth LWT, Mameren H van, Prince CJA, Scherpbier AJA, Vleuten CPM van der. Testing in context: the effects of clinical context on the results of an anatomy test. *Dutch Journal of Medical Education* 2002;21(6): 259-266.)